

*Tamás Szirányi – László Havasi*

## **Observation on Earth and from the Sky How Can the Events Embedded in our Environment Be Made into Useful Data?**

**Tamás Szirányi**, Research Fellow of the Budapest University of Technology and Economics; Head of Laboratory of the Institute for Computer Science and Control, Machine Perception Research Laboratory of the Hungarian Academy of Sciences

**László Rajmund Havasi**, Research Fellow of the Institute for Computer Science and Control, Machine Perception Research Laboratory of the Hungarian Academy of Sciences

### **Abstract**

Our safety is based on the continuous monitoring and evaluation of the measured world. The detected images may come from surveillance video systems, on-board systems of vehicles or even from satellites. With the development of technology, the types of sensor devices, the perceived contents of the physical signals, or the placement has become diverse: observation is possible from virtually anywhere at any time. We require new mathematical solutions or information management procedures for the analysis of information perceived from different observation points at different time instants, and comparing and evaluating data concatenated in space or time, interpreting events or situations as complex problems. In this article we briefly describe the typical measurement data and equipment storehouse, indicating what tasks and opportunities arise in the understanding of the dense world of detected signals.

*Keywords:* observation, recognition, sensor-networks, video camera system, area surveillance, Earth observation, remote sensing, biometrics

### **Introduction**

Video surveillance systems first became ubiquitous in large cities, and were followed by the security systems of banks and certain public buildings. At the time, cameras were actively monitored by operators. If there were multiple cameras installed, the security staff checked the video feed switching between cameras on a split screen. However, when there was a camera at every corner, most of them working in scanning mode, the number of video moni-

tor personnel rocketed beyond the manageable threshold. Later, security cameras became an important fixture of houses and apartment blocks. Most people no longer bother to look at them: the knowledge that acts of crime can quickly be investigated and solved by reviewing the recorded material makes them feel secure enough.

Placing cameras on public passenger transport vehicles is meant to have the same effect.

However, with the rise of terrorism, prevention and constant monitoring became important again. In order to overcome the difficulties arising from excessive workload and lapse of concentration, at least the filtering aspect of receiving video streams should be automated and relevant information should be provided to the operators, which means that information should be filtered out during pre-processing. Recently artificial vision functions of self-driving vehicles came into the limelight and automated aircraft were added to the military arsenal for aerial reconnaissance and intervention. These use several sensors some of which might even operate on the basis of different principles, and the goal is to fuse their data and extract (semantic) information of higher value from them.

These new sensing technologies gave rise to new problems. It is not just law enforcement and area defence that employs new devices, but criminals too: detection and elimination of UAVs equipped with spy cameras or posing a terror threat has become important. We only have a few years to develop countermeasures against the emerging new technologies while technical progress in this area has accelerated extremely.

## **Tasks**

### **Area defence**

Fences and walls are meant to keep people out, but they only provide physically limited defence. Active surveillance devices allow for constant monitoring and alerting operators to suspicious movement. All surveillance methods can be evaded by, for example, moving skilfully, hiding, blinding the sensor, or disrupting the network. The simpler the system is, the greater the danger that a well-prepared criminal will circumvent it. If we use multiple sensors (optical, infra, PTZ motion detection, wide-angle, and full panorama), then we have a chance to thwart trickery. Automated alerts can be extremely important as they can help security guards whose attention has lapsed immensely.

### **Proactive anti-terrorist surveillance**

The movement and appearance of terrorists differ from the way ordinary criminal intruders act. Beyond a certain point they do not aim to conceal themselves; rather, they strive to act quickly. In order to frustrate their efforts, an extended alarm zone must be created around the facility, and special behaviour patterns typical of terrorists must be detected in time. Proactivity is particularly important so that preparations may be detected as early as the preparatory/reconnaissance stage. These clues that warrant suspicion might be insignificant in themselves, but multiple factors may justify triggering an alarm. Such a clue could be

when a van pulls over, but nobody gets out of it, and a motorist appears who seems to be surveying the area.

## Disaster Management

In the event of disasters that occur or have already occurred (e.g. gas explosion, flood), the target, the participants and the environment are well defined, but there is no enemy to overcome, only a task that must be tackled. In such a case, speed is crucial so you can deploy the most capable sensors near the target as soon as possible: robot sensors, UAV vehicles with on-board cameras and chemical detectors, and fixed cameras to detect changes. Fusion of various data sources, for example, comparison of daily satellite images with data from cameras aboard UAVs on mission can be important to that end.

### Tools to detect secure environments

The tools of visual detection are mostly high-resolution colour and infra cameras as well as depth sensors that can (also) detect spatial distance, and the hybrid versions of these devices.

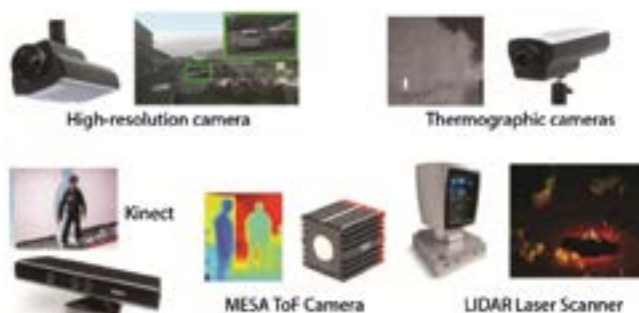


Figure 1

*Imaging elements of security systems: cameras and depth sensors (3D)*

*Source: The author's own contribution*

## Cameras

Today camera resolution is in the 30 to 150 MP (megapixel) range. If it is paired with high quality optical elements, then its true resolution can show very fine details. A good zoom camera can display views from wide-angle images to finely detailed close-ups. It is important to note, however, that if the resolution of the image sensor matrix does not match that of the optical elements, we will get a poor quality or even corrupted image.

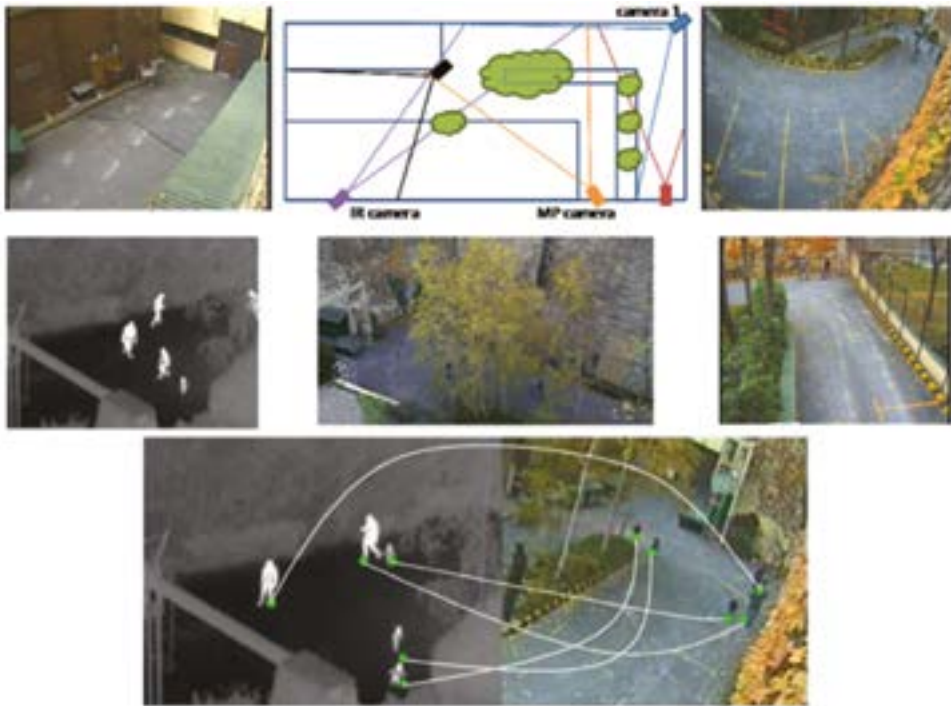


Figure 2

*Camera images taken at various locations<sup>1</sup> (correlation between moving objects in various views is established through the use of 3D calibration procedures)*

*Source: KISS–SZIRÁNYI, 2013; HAVASI et al., 2014*

## Depth Sensors

3D (or, rather, 2.5D depth) cameras are already available in the market. As the technology becomes more sophisticated, they will be used in large numbers, replacing/supplementing stereo technology. Their scale ranges from toys to expensive laser scanners.

### *Kinect*

The computer gaming industry created the Kinect<sup>2</sup> sensors that can recognize hand gestures, allowing for virtual control over any movement. This device was essentially conceived as a toy, but it grew out of this role, and it is now used in, for example, medical diagnostics. Its range is limited and it only works indoors, just like many other devices designed for similar uses.

<sup>1</sup> ProActive project (EU FP7) of HAS ICSC, [http://cordis.europa.eu/project/rcn/103500\\_en.html](http://cordis.europa.eu/project/rcn/103500_en.html)

<sup>2</sup> <https://developer.microsoft.com/en-us/windows/kinect>

### ToF

In time, however, cameras<sup>3</sup> that, in addition to normal images, create depth images too have been released. The principle of their operation is based on assigning a *time-of-flight* value to each pixel of the refracted reference radiation, which can be used to create a depth map. Theoretically these cameras can be used outdoors, but their accuracy (about 1 cm) and their sensitivity still need to be developed further. However, they can provide a high frame rate (about 160 fps).

### Laser scanners

In LIDAR<sup>4</sup> devices, a laser beam scans the environment. The laser light typically has a wavelength in the 600 to 1000 nm range. With regard to the mode of operation, they have two basic types:

- A static or architectural LIDAR: they are placed at a certain location, and then a rotating mirror system scans the environment within a radius of 200 to 300m, and provides results about the surrounding 3D structure (point cloud) with an accuracy of millimetres. The point clouds can be numerically evaluated: objects and surfaces can be identified, and meshes can be laid over the surface. Repeating the measurement after a long time the two measurement results can be matched in 3D, even if they were taken at two different locations. This allows for finding differences, for example the tilt of columns and the displacement of walls can be accurately measured. This feature is also used for forensic investigation of accidents or documenting archaeological excavations.
- Dynamic LIDAR devices: these are laser scanners mostly made for self-driving vehicles (type Velodyne<sup>5</sup>), which create mappings at a rate of 1 million spatial points/s, using multiple beams vertically while rotating horizontally with a sampling rate of 20 rotations/sec. They can assist the self-driving functions of cars in 3D, map spatial street views, recognize 3D shapes (Rózsa-Szirányi, 2016), and measure real-time spatial movements.
- A variation of the first two categories where the second scanning direction of a LIDAR that is scanning in single or multiple lines is determined according to the speed of the carrier vehicle for scanning, for example, tunnels and walls of buildings.

---

<sup>3</sup> <http://hptg.com/industrial/>

<sup>4</sup> Laser radar.

<sup>5</sup> <http://velodynelidar.com/>

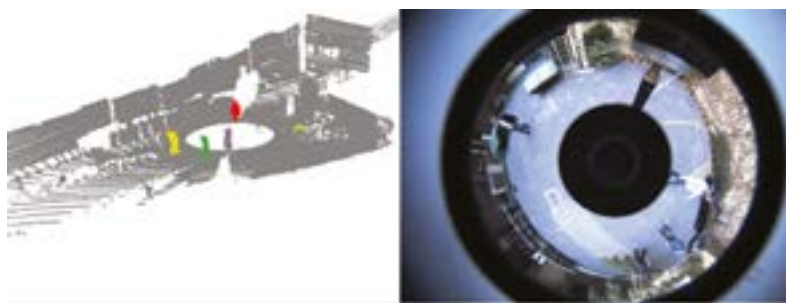


Figure 3

*3D point cloud of a LIDAR scan<sup>6</sup> (on the left) and a 360° panorama image captured through the lens of a PAL<sup>7</sup> panorama camera, taken of the same location (on the right)*

*Source: Author's own images*

### Sensors installed on UAVs

The popularity of unmanned aerial vehicles is growing, and their capabilities and on-board equipment are progressing rapidly. Their flight time keeps increasing, while the capabilities of the devices installed on them are getting better and better, with a parallel decrease in weight.

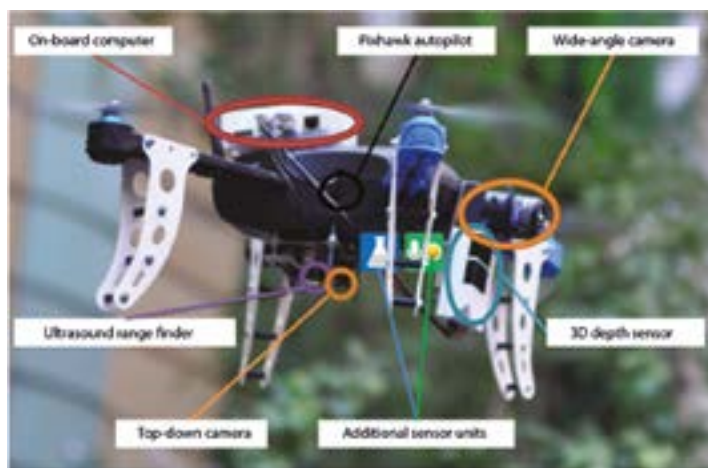


Figure 4

*A quadcopter equipped with cameras and other sensors – these vehicles are becoming safer (automated return-to-home and landing unit) and their flight distance has also been increasing*

*Source: Illustration of András Majdik (2015)*

<sup>6</sup> Recognized pedestrians are coloured (Csaba Benedek and his colleagues, i4D project, HAS ICSC, Budapest).

<sup>7</sup> Invention of Pál Greguss (BME, 1983); it was purchased by the Japanese company Tateyama, and was manufactured by Sony until recently. Patent has expired.

## Aerial and satellite photos

The quality of images downloaded from satellites and their access time are improving. Various images from the Sentinel satellites of ESA can be accessed from a given area several times a day, and at no cost.<sup>8</sup> In this way, security surveillance and monitoring is available from space. Combined with aerial photos taken by UAVs, satellite pictures can provide us with an up-to-date, inexpensive and easy-to-access source of images. In addition to security management, the other beneficiaries of this service are disaster management, agriculture, regulatory supervision (built-up areas) and environmental protection (e.g. detection of massive deforestation).



Figure 5

*Automated detection of built-up areas from aerial photos (detection of populated areas is shown on the right side)*

Source: KOVÁCS–SZIRÁNYI, 2013

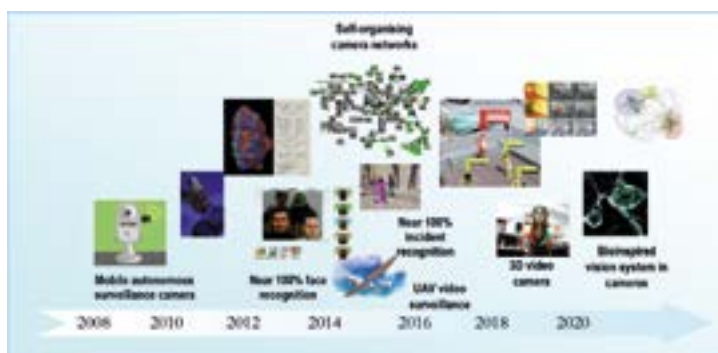


Figure 6

*Development of the individual technological elements (the years are conservative estimates, the actual progress may be faster in the upcoming years)*

Source: PETRÁS, 2008

<sup>8</sup> <https://sentinel.esa.int/web/sentinel/home>

## **Special data and how to manage them**

Depending on how the data was generated and the type of target area, there are various rules applicable to the storage and processing of captured images. Sensitive data may include biometric properties (e.g. face, fingerprint), but in certain cases, even aerial photos. Data management should take into account privacy rights and information values. Often, if the data have already been interpreted and the essential properties have been extracted, the original data are no longer needed (they can be deleted), and thus a large database can be created by processing sensitive data, but without ethical/legal ramifications. Data generated by the analysis of image data may often contain image sets and camera images taken from multiple views. In such cases, linking the individual images together requires the application of the efficient tools of multiple view image geometry and motion analysis, which, in turn, may necessitate adding extra computing capacity and a linked database. This might also entail the use of data processing centres, including, for example, cloud architectures where data security is of particular concern.

## **Biometric analysis**

Certain biometric properties such as facial image, fingerprint, and signature have become legal means of providing proof of identity. Other characteristics such as iris, retina, and palm scans, and chemical properties also play a major role in security technology. But there are other properties, for example, the detection of gait (HAVASI et al., 2007), body structure, facial expressions, hand gestures, that can facilitate the detection of a limited group of persons or, rather, the tracking of their motion as they pass through several sensors' detection areas. The security of biometric data is very problematic, and filtering out fraudulent attempts requires complex procedures (e.g. using fingerprints together with thermal images of the hand). It is required in this professional field that data captured for authentication (e.g. for an access control system) should be incomplete, which means that these system should only extract and store as much information as absolutely necessary and automatic merging of various databases should not be possible.

## **Information technology background**

Storing the increasing number of differently sized images, indexing, processing and searching them in databases, and comparing file content (e.g. to detect changes between different points in time) is becoming more and more difficult. Working with large database falls within the domain of Big Data, while processing images and recognizing objects (VARGA-SZIRÁNYI, 2016) are tasks that require machine learning (e.g. deep learning) and the processing of large databases. Processing data from each successive year necessitates the use of new mathematical procedures that rely on stochastic optimization (SZIRÁNYI-SHADAYDEH, 2014). Calibration of various tools, co-registration of images (VARGA et al., 2015), and fixing distortions is impossible without serious mathematical capabilities; research and development in this area requires constant efforts. In order to promote human management and



database look-up activities, the automated filtering and learning procedures of large image databases present the greatest challenge in the area of artificial intelligence.

## Legal background

The use of cameras in public spaces is restricted by legal provisions that were prepared with the involvement of the authors (PETRÁS et al., 2008). Today it is only natural that image capture devices record events in public areas, including vehicle related incidents. Retrieving these data is only permitted if the applicable authorization is obtained, and thus such data do not violate privacy. We have already grown accustomed to the presence of cameras everywhere – all for the sake of our protection. Taking aerial photos by using UAVs is subject to legal restrictions; banning such vehicles from private and protected public area is technically feasible (e.g. by integrating the solution into the navigation system's database). As in the case of mapping and data storage regulations pertaining to public camera systems, publication of recordings taken in public areas is also subject to legal restrictions, and it is regulated accordingly.

## References

- HAVASI László – KISS Attila – SPÓRÁS László – SZIRÁNYI Tamás (2014): Calibrationless sensor fusion using linear optimization for depth matching. *LNCS*, Vol. 8466, Springer, 158–170.
- HAVASI László – SZIRÁNYI Tamás (2016): Multi-View Security and Surveillance at MTA SZTAKI. *ERCIM NEWS*, Issue 106. 55–56.
- HAVASI László – SZLÁVIK Zoltán – SZIRÁNYI Tamás (2007): Detection of Gait Characteristics for Scene Registration in Video Surveillance System. *IEEE Tr. Image Processing*, Vol. 16, No. 2. 503–510.
- KISS Ákos – SZIRÁNYI Tamás (2013): Localizing People in Multi-view Environment Using Height Map Reconstruction in Real-Time. *Pattern Recognition Letters*, Vol. 34, No. 16. 2135–2143.
- KOVÁCS Andrea – SZIRÁNYI Tamás (2013): Improved Harris feature point set for orientation sensitive urban area detection in aerial images. *IEEE Geoscience and Remote Sensing Letters*, Vol. 10, No. 4. 796–800.
- PETRÁS István et al. (2008): *Út a térfigyelő rendszerek stratégiájának elkészítéséhez*. Közterületi-Térfelügyeleti Tanácsadó Testület, Budapest.
- RÓZSA Zoltán – SZIRÁNYI Tamás (2016): Exploring in partial views: Prediction of 3D shapes from partial scans. *12<sup>th</sup> IEEE International Conference on Control & Automation, ICCA'16*. IEEE, Kathmandu.
- SZIRÁNYI Tamás – SHADAYDEH, Maha (2014): Segmentation of remote sensing images using similarity measure based fusion-MRF model. *IEEE Geoscience and Remote Sensing Letters*, Vol. 11, No. 9. 1544–1548.
- VARGA Domonkos – SZIRÁNYI Tamás (2016): Robust real-time pedestrian detection in surveillance videos. *Journal of Ambient Intelligence and Humanized Computing*, Vol. 8, No. 1. 213–221.

VARGA Domonkos – SZIRÁNYI Tamás – KISS Attila – SPÓRÁS László – HAVASI László (2015): A multi-view pedestrian tracking method in an uncalibrated camera network. *International Conference on Computer Vision: ICCV*. IEEE, Santiago de Chile.